African Journal of Biotechnology Vol. 11(9), pp. 2303-2312, 31 January, 2012
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB11.2648

Full Length Research Paper

Rapid and non-destructive discrimination of tea varieties by near infrared diffuse reflection spectroscopy coupled with classification and regression trees

Shi-Miao Tan[1], Rui-Min Luo[1], Yan-Ping Zhou[1*], Hong Gong[2] and Ze Tan[3]

[1]Key Laboratory of Pesticide and Chemical Biology of Ministry of Education, College of Chemistry, Central China Normal University, Wuhan 430079, P. R. China.
[2]Economics and Management School, Wuhan University, Wuhan 430072, P. R. China.
[3]State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, P. R. China.

Accepted 28 October, 2011

The current study attempted to rapidly and non-destructively discriminate the diverse varieties of tea (that is, Biluochun, Longjing, Maojian, Qihong, Tieguanyin, and Yinzhen) via utilizing near infrared (NIR) diffuse reflectance spectroscopy coupled with pattern recognition strategies. Before the recognition analysis, the original NIR spectra were pre-processed by second derivative treatment followed by informative wavenumber interval location. And then, non-linearity detection and outlier diagnosis were performed. When pattern recognition referred, principal component analysis (PCA) was firstly applied to ascertain the discrimination possibility with the NIR spectra. Classification and regression trees (CART), compared with linear discriminant analysis (LDA), and partial squares-discriminant analysis (PLS-DA), was then employed for establishing the discrimination rule. Experimental results showed that the tea quality could be accurately, rapidly, and non-invasively identified via NIR spectroscopy coupled with CART.

Key words: Near infrared diffuse reflection spectroscopy, classification and regression trees, and tea variety discrimination.

INTRODUCTION

Tea, manufactured from the leaves of the *C. sinensis* plant, is the most widely consumed drink throughout the world after water. This may be in relation to the beneficial medicinal properties of tea (Nakachi et al., 2000; Setiawan et al., 2001; Fujiki et al., 2001; Yang et al., 2002), such as, the protective effects against coronary heart diseases (Nakachi et al., 2000), chronic gastritis (Setiawan et al., 2001), and several caners (Fujiki et al., 2001). Up to the present, many commercial tea varieties have sprung up in the market, holding different botanicals and quality. Such differences are commercially evaluated and appreciated by the consumers, usually being the important factors to determine the price of the commercial tea products. Nowadays, food safety and authenticity have attracted considerable attention worldwide. Nevertheless, driven from the illegal commercial benefits, the phenomena of shoddy and adulteration could be found everywhere. These seriously destroy the consumer benefits, and also make against the maintenance of the tea brand. Moreover, with the development of international trade and the adjustment of human diet framework, the higher

*Corresponding author. E-mail: hgzyp2005@yahoo.com.cn. Tel: +86-15872406428. Fax: 86-27 67867141.

Abbreviations: CART, Classification and regression trees; **PCA,** principal component analysis; **PLS-DA,** partial least square-discriminant analysis; **LDA,** linear discriminant analysis; **NIR,** Near infrared**; MCCP,** minimal cost-complexity pruning.

tea quality is required than before. For instance, some criteria for tea quality factors have been specified by several national and international authorities (Fujiki, et al., 2001). In general, caffine, free amino acids, and polyphenols have been considered as the important quality components of tea. The contents and proportion of these three quality components are mainly responsible for the flavor of the tea brew, being different for diverse tea varieties more or less. Hence, it is of great importance to distinguish the inter-variety difference of tea and control the tea quality. This will help promote the intrinsic quality control of tea, prevent infringement and also beat fake and shoddy products in the open market.

Traditionally, the discrimination of tea varieties has been executed by using wet chemical analysis tools to detect the disparities of the principal quality factors among the different tea varieties. The involved wet chemical methods are precise but time consuming, laborious and invasive, including capillary electrophoresis, gas chromatography, high performance liquid chromatography (HPLC), and colorimetric measurements. Near infrared (NIR) spectroscopy, as a rapid, simple and non-destructive technique, can be a better alternative to the wet chemical tools, examples of which are associated with its extensive applications in agricultural (Alain et al., 2008), manufacturing (Jens and Jacques, 2008; Shao et al., 2010), pharmaceutical (Christoffer et al., 2005; Sun et al., 2008) and food industries (Liu et al., 2000; Šašić and Ozaki, 2001; Jing et al., 2010). Since Hall et al. (1988) utilized NIR spectroscopy rapidly and non-invasively in predicting the quality of black tea, the application of NIR spectroscopy for tea quality assessment has been the subjects of many investigations (Schulz et al., 1999; Luypaert et al., 2003; Zhang et al., 2004; Chen et al., 2006; Chen et al., 2008; He et al., 2007). For instance, Schulz has employed NIR coupled with partial least squares regression to simultaneously predict the alkaloids and phenolic substances in green tea leaves (Schulz et al., 1999). Massart et al. have tested the possibility of NIR in quality analysis of green tea, and then applied NIR for the total antioxidant capacity analysis in green tea (Luypaert et al., 2003; Zhang et al., 2004). In China, Chen et al. (2006, 2008) and He et al. (2007) have also dedicated their studies to tea quality evaluation via NIR. However, most of these publications have focused on the quantitative analysis of tea (Chen et al., 2006; 2008), especially green tea. Few studies dealing with tea variety discrimination have been reported until now (He et al., 2007).

In this study, tea variety discrimination, exemplified as Biluochun, Longjing, Maojian, Qihong, Tieguanyin, and Yinzhens has been carried out by using NIR diffuse reflectance spectroscopy coupled with pattern recognition techniques. The NIR spectra obtained from six tea varieties have been combined as one data set for the discrimination analysis. Such a data set may present some inherent characteristics, viz., co-linearity, non-linearity, heterogeneity, or existence of outliers. Among the existed investigations regarding with the tea quality analysis, attention has been almost not paid on the data quality control. Hereon, before pattern recognition, the data quality control has been carried out, including outlier and non-linearity detection. This is beneficial for further selection of a suitable pattern recognition tool so as to establish an efficacious discrimination rule. In addition, classification and regression trees (CART) has been considered to implement the discrimination task. This is attributed to the fact that CART holds promising modeling performance and many attractive features, including simplicity, interpretability, high capacity in handling large data set and modeling non-linearity, suitability for multi-class issue, no assumption regarding the data distribution and immunity to outliers, co-llinearity and heteroscedasticity. Simultaneously, as comparisons, principal component analysis (PCA), partial least square-discriminant analysis (PLS-DA), and linear discriminant analysis (LDA) have also been investigated. Results have shown that the inter-variety difference of tea can be well discriminated by NIR spectroscopy coupled with CART.

MATERIALS AND METHODS

Sample preparation

Six varieties of tea leaf samples, purchased from the supermarket, are originated from different provinces in China. The tea leaves were crushed via a beater mill and sieved with a 0.5 mm screen. And this sieved tea powder was used for the further NIR spectral measurement. The tea information is summarized in Table 1, including the categories, varieties, origins and numbers.

Acquisition of NIR data

The NIR spectra for the powdered tea samples were collected on an Antaris II NIR spectrometer (Thermo Electron Co., USA) in the reflectance mode. This NIR spectrometer is furnished with a quartz sample cup, an integrating sphere and an indium gallium arsenide (InGaAs) detector. Under a steady level of temperature and humidity, the NIR spectral measurement was performed with the spectral range and the resolution as 4000 to 10000 cm^{-1} and 8 cm^{-1}, respectively. In addition, a total of 64 scans were accumulated per measurement. The background was collected at the beginning of this experiment and then after every one hour.

The obtained 300 spectra were in turn collected from 50 Biluochun, 50 Longjing, 50 Maojian, 50 Qihong, 50 Tieguanyin and 50 Yinzhen samples. For making use of LDA, PLS-DA, and CART to perform and evaluate the discrimination task, the whole data set was randomly split into a calibration set of 176 samples and a prediction set of 124 samples. The calibration set was composed of 25

Table 1. Categories, varieties, numbers and origins of tea samples.

Tea variety	Sample number	Tea origin	Tea category
Biluochun	50	Jiangsu	Green tea
Longjing	50	Zhejiang	Green tea
Maojian	50	Henan	Green tea
Qihong	50	Anhui	Black tea
Tieguanyin	50	Fujian	Oolong
Yinzhen	50	Hunan	Yellow tea

Biluochun, 30 Longjing, 33 Maojing, 32 Qihong, 26 Tieguanyin and 30 Yinzhen tea samples. For the prediction set, the remainders were embodied.

Classification and regression trees

CART (Breiman et al., 1984), as a binary tree representation, is able to describe the relationship between the dependent and independent variables with high flexibility and sufficient accuracy. The dependent variable can be either numerical or categorical, respectively, resulting in regression or classification tree. An extended depiction on CART can be referred to Breiman et al's excellent book, *Classification and Regression Trees*. Here, since CART was used for classification task, only a concise description of the classification tree is presented.

Generally, the configuration of CART consists of three basic steps. Firstly, the largest tree is grown by applying greedy recursive partitioning. Recursive partitioning is conducted in top-down fashion, starting from the root node containing the entire calibration samples until each node reaches complete homogeneity or a user- specified minimal sample number and becomes a terminal or leaf node. Then, on the basis of the minimal cost-complexity pruning (MCCP) criterion (Breiman et al., 1984), the largest tree is pruned to yield a sequence of nested sub trees. Ultimately, from such nested sub trees, the final appropriately-fit CART is selected in terms of its best prediction accuracy either gained by the cross validation method or pruning set technique (Breiman et al., 1984). Once the final tree is gained, some imminent node information, that is the splitting rule and node output is endowed. A prediction of the variety of an unseen sample from its given NIR spectroscopy is implemented by traversing the tree till a leaf node is reached, and this leaf node output acts as the predicted variety.

Software

The algorithms used in the current study were written in Matlab environment and run on a personal computer with the processor as Intel Pentium (R) Dual-Core CPU E6300 @ 2.80 GHz 2.79 GHz and the RAM as 2GB.

RESULTS AND DISCUSSION

Figure 1a depicts the raw NIR spectra for the powdered tea samples, showing three clusters of absorption bands. The band cluster ranging from 4000 to 5000 cm^{-1} might be attributed to the second overtone of C-H deformation mode, the aggregations of the O-H and N-H combination modes. The two bands from 5200 to 7200 cm^{-1} possibly arises from the C-H stretching overtone and the O-H and N-H stretching overtones, respectively. The weak absorption peak around 8500 cm^{-1} is due to the second overtone of C-H stretching mode. All of these absorbance bands may be caused by the multi-ingredients of tea, such as polyphenols, alkaloids, proteins, amino acids, and some aromatic compounds. It can also be seen that the original NIR spectra of different samples have considerable baseline drifts. This might be due to the fact that the powdered samples show slightly varying characteristics in the reflection and the scattering of the incident NIR light. The baseline drifts could be effectively eliminated by the second derivative treatment, as indicated in Figure 1b. In addition, it was shown that the wavenumber interval selection procedure could eliminate the extra variability generated by non-composition related factors such as perturbations in the experimental conditions and the physical properties of samples, thereby offering ameliorated performance for multivariate spectral analysis (Jiang et al., 2002). Consequently, the derivative NIR spectra were analyzed using the moving window partial least squares regression method (Jiang et al., 2002). This method ascertained several informative spectral regions that is, 4040 to 4180, 4570 to 4690, 5130 to 5380, 5830 to 5940, 6790 to 6890, 6970 to 7310, and 7510 to 7620 cm^{-1}. The pre-processed spectra were used for the ultimate tea variety identification analysis.

Since such a data set may present some inherent attributes, for example, non-linearity or outliers, data quality control before recognition analysis was performed. It is of great advantage for further selecting the suitable recognition method. Runs test method (Centner et al., 1998) was applied to test whether non-linearity exists or not and quantify the extent if non-linearity was present. For this data set, the presence of serious non-linearity was proven by the runs test that yielded a statistical value of -16.2451 whose absolute value is much larger than

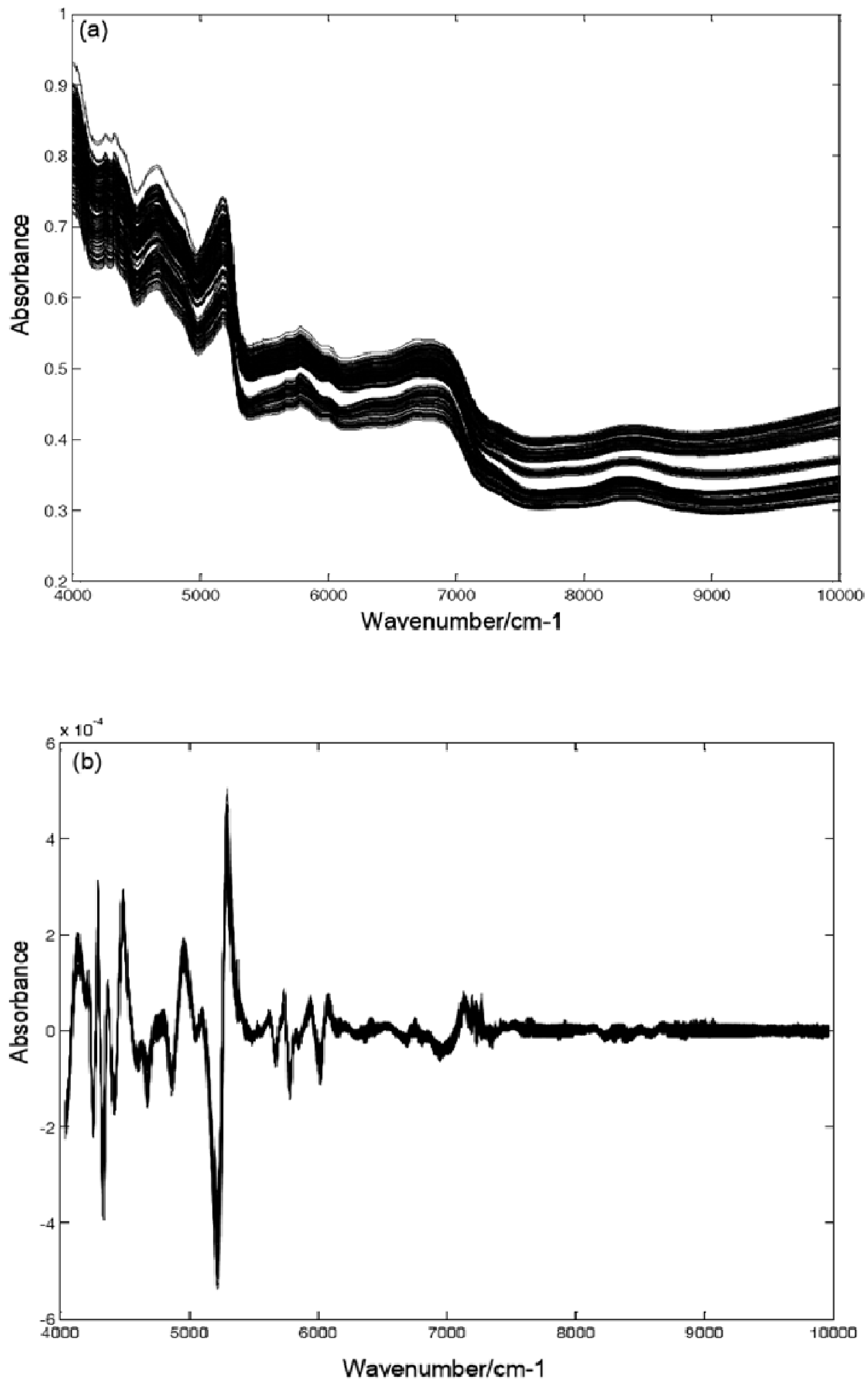

Figure 1. (a) Raw NIR spectra of six varieties of tea powder. (b) Second-derivatives of the raw NIR spectra.

the critical value of 1.96 (Centner et al., 1998). For the outlier diagnosis, the Cook's squared distance ($CD^2_{(i)}$) method was used (Massart et al., 1997). The large value of $CD^2_{(i)}$ indicates that the *i*th sample plays a considerable influence on the regression estimator. Generally, $CD^2_{(i)}$ = 1 is considered to be large (Massart et al., 1997). The results of the outlier diagnosis are depicted in Figure 2, indicating the absence of outlier.

As for the recognition analysis, PCA was firstly used for

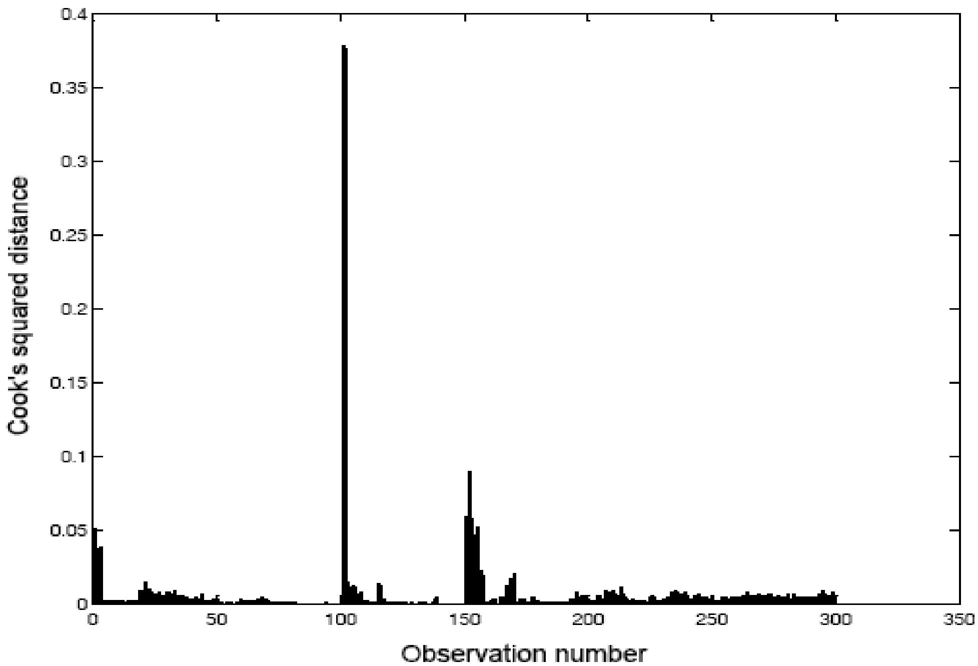

Figure 2. $CD_{(i)}^{2}$ Values obtained by Cook's squared distance method.

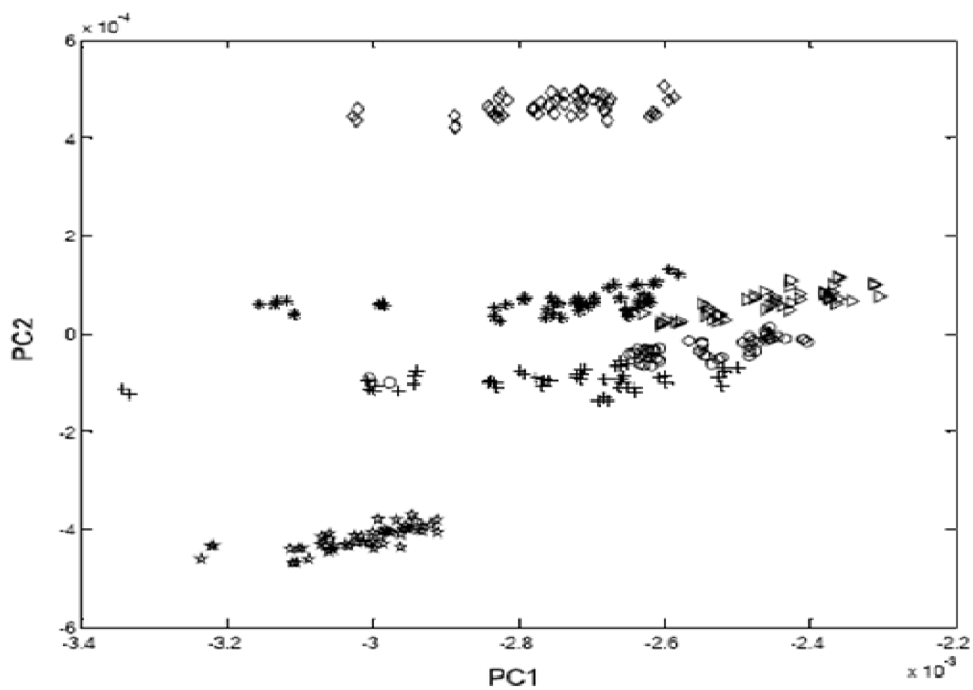

Figure 3. Score plot by the top two principal components for 300 powdered tea samples, where symbols ○ (circle), ▷ (triangle-right), + (plus), * (star), ◇ (diamond) and ☆ (pentagram) represent Biluochun, Longjing, Maojian, Qihong, Tieguanyin and Yinzhen tea samples, respectively.

exploring the cluster tendency. The score plot by the top two principal components is shown in Figure 3. It can be observed from Figure 3 that although some points overlap each other, the obvious cluster tendency occurs. This good cluster tendency may be due to the fact that the diverse varieties of tea are in possession of considerable differences in their botanical, genetic, and agronomical characteristics, and especially the treatment processes and the original regions. In addition, via a visual inspection of Figure 3, one can obtain that three kinds of green tea lie adjacent to each other. Qihong, a category of black tea, lies adjacent to the three varieties of green tea and only makes a little super-position with Longjing as shown in Figure 3. Tieguanyin (oolong) and Yinzhen (yellow tea) are separated fully from each other and clearly differentiated from the other four varieties of tea without overlapping. The distinctness of inner chemical compositions among the diverse categories of tea leads to the occurrence of such a phenomenon, mainly resulting from the dissimilarities of manufacturing procedures among these four categories of tea. For green tea, the tea leaves are only roasted but not fermented before being dried, containing more of the simple flavonoids called catechins. The tea leaves for yellow tea are slightly fermented via a step called “sealed yellowing” before being dried. Oolong and black tea, respectively, are semi-fermented and fully fermented tea. The fermentation is an oxidation procedure converting the catechins into the theaflavins and thearubigins. Hence, different categories of tea present various chemical characteristics. All of these indicated that the minor NIR spectroscopy difference can provide enough information for the further tea variety discrimination analysis.

In the current study, LDA was firstly used for the tea variety discrimination. LDA is a well-known technique for dimensionality reduction and feature extraction. In LDA, several orthogonal latent variables were searched to represent the original feature space, holding the maximum between-group variance compared with the within-group variance of all the groups. The essence of this method is to extract a low-dimensional subspace with strong discriminant ability from the original feature space; then the scores of the first 2 or 3 latent variables are plotted. In this study, the score plots by LDA are demonstrated in Figure 4. It can be seen that six varieties of tea in the calibration set are identified accurately while those in the prediction set suffer from serious overlapping and confusion, indicating that LDA presents good learning performance but less effectiveness to the unknown samples. This may result from the fact that the data set possesses strong non-linearity, and in which, the variable numbers are much larger than the sample ones.

PLS-DA, as a linear technique, was also employed to relate the measured NIR signals to the variety memberships of tea samples. PLS-DA aims to find the variables and directions in multivariate space which discriminate the known classes in the calibration set. In PLS-DA, the variety memberships of samples are coded as a dummy matrix containing as many columns as the known classes in the calibration set. Each sample is assigned a value of 1 or 0 depending on whether it belongs to the class represented by that column or not. In the present study, the dummy codes for Biluochun, Longjing, Maojian, Qihong, Tieguanyin, and Yinzhen can be represented as follows: (1, 0, 0, 0, 0, 0), (0, 1, 0, 0, 0, 0), (0, 0, 1, 0, 0, 0), (0, 0, 0, 1, 0, 0), (0, 0, 0, 0, 1, 0) and (0, 0, 0, 0, 0, 1). Calibration was then carried out by regressing the optical data on the dummy variables. The estimated dummy matrices obtained by PLS-DA for the calibration and prediction sets are shown in Figures 5a and b, respectively. The latent variable for PLS-DA is 3 optimized on cross validation method. Six varieties of tea samples can be recognized via locating the maximal dummy codes of the samples, respectively. For example, as shown in Figure 5a, the maximal values of the estimated dummy codes for the first two samples locate in the sixth subplot (that is, t6), indicating that these two samples essentially belonging to Biluochun are misclassified into Yinzhen. Via an aborative inspection of Figure 5, one can obtain that although majority of tea samples are identified accurately, however, quite a few are misclassified. Table 2 lists the classifi-cation results from which one can obtain that the total recognition rates (RR) for the calibration and prediction sets are 89 and 83%, respectively. PLS-DA provided the RRs of 48 and 79%, respectively, for the varieties of Biluochun and Maojian in the calibration set. The RRs for these two varieties in the prediction set were 32 and 76%. For the remainder four varieties, PLS-DA provided the RRs of 100% in both the calibration and prediction sets. It seems that PLS-DA recognizes the diverse categories of tea with relatively satisfactory results but provides relatively poor discrimination ability among green tea. These results are tally with those obtained by PCA. This may be due to the component similarities among green tea and the deficiency of PLS-DA in calibra-ting the data set possessing complex and unknown non-linearity.

Finally, CART was invoked to more effectively execute the variety recognition task. During CART configuration, the minimal node size was specified as 5, that is, the node covering less than 5 samples could not be split further and was specified as the leaf node. The appropriately-fit CART was identified in terms of its best prediction accuracy gained by the leave-one-out cross validation method. The tree obtained by CART is shown in Figure 6, furnished with some imminent node information, that is, the splitting rule and node output. A prediction of the variety of an unseen sample from its given NIR spectroscopy is implemented by traversing the tree till a leaf node is reached, and this leaf node output acts as the predicted variety. For instance, if an unknown sample is located in node 14 via traversing the tree, this sample is predicted as Maojian. In CART, no dummy coding for the memberships is needed and the variety identification is very intuitionistic. The recognition results of CART are also enumerated in Table 2. By using CART, the total RRs for the calibration and prediction sets were improved from 89 to 99% and 83 to 94% by PLS-DA, respectively. When compared with PLS-DA, CART yielded much higher RRs for Biluochun and Maojian in

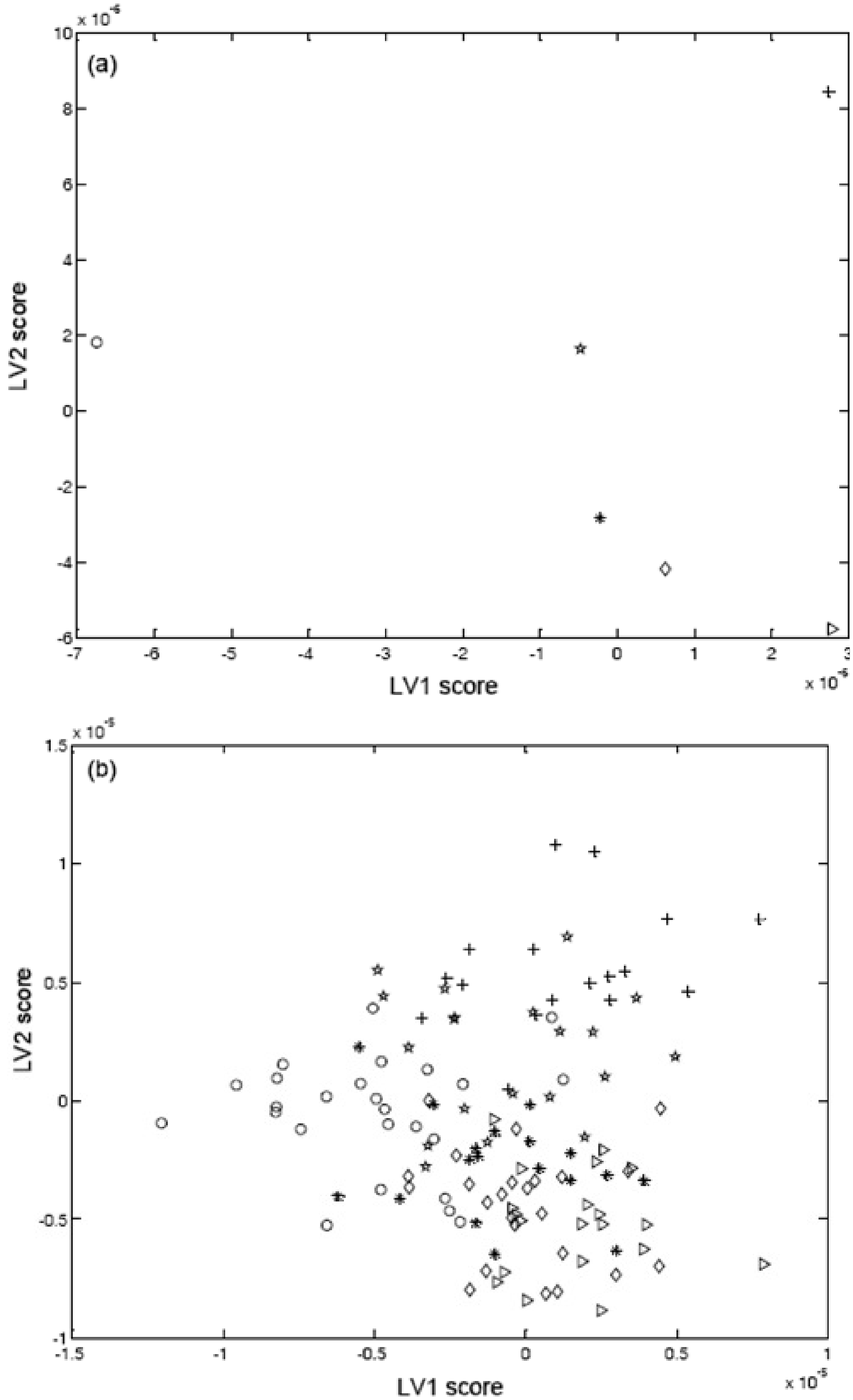

Figure 4. Score plots by the first and second latent variables for (a) 176 calibration tea samples and (b) 124 prediction tea samples, where ○ represents 25 calibration and 25 prediction Biluochun tea samples; ▷ stands for 30 calibration and 20 prediction Longjing tea samples; + refers to 33 calibration and 17 prediction Maojian tea samples; * stands for 32 calibration and 18 prediction Qihong tea samples; ◇ refers to 26 calibration and 24 prediction Tieguanyin tea samples; and ☆ represents 30 calibration and 20 prediction Yinzhen tea samples.

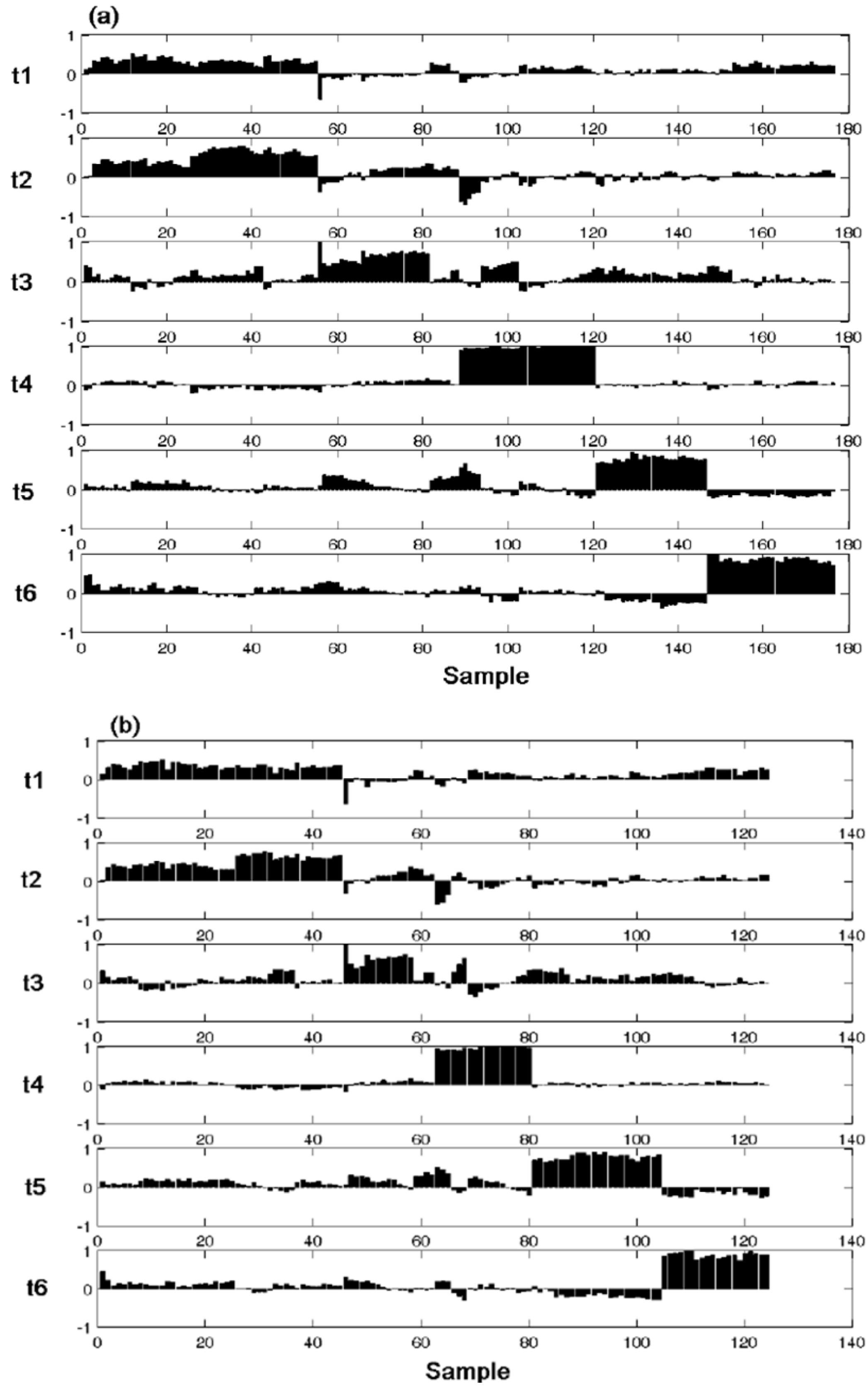

Figure 5. The estimated dummy matrix by PLS-DA refers to the variety memberships of the samples in the (a) calibration set and (b) prediction set. Thereinto, the six subplots (that is, t1, t2, t3, t4, t5, and t6) in turn represent six columns in the estimated dummy matrix versus sample number.

Table 2. Classification results using CART compared with those obtained by PLS-DA.

Data set	Variety	PLS-DA			CART		
		Number of misclassified sample	RR (%)	Total RR (%)	Number of misclassified sample	RR (%)	Total RR (%)
Calibration set (including 176 samples)	BLC	13	48	89	0	100	99
	LJ	0	100		1	97	
	MJ	7	79		0	100	
	QH	0	100		0	100	
	TGY	0	100		1	96	
	YZ	0	100		0	100	
Prediction set (including 124 samples)	BLC	17	32	83	3	88	94
	LJ	0	100		3	85	
	MJ	4	76		1	94	
	QH	0	100		0	100	
	TGY	0	100		1	96	
	YZ	0	100		0	100	

BLC, LJ, MJ, QH, TGY and YZ, respectively, represent Biluochun, Longjing, Maojian, Qihong, Tieguanyin and Yinzhen tea samples. RR is the abbreviation of recognition rate.

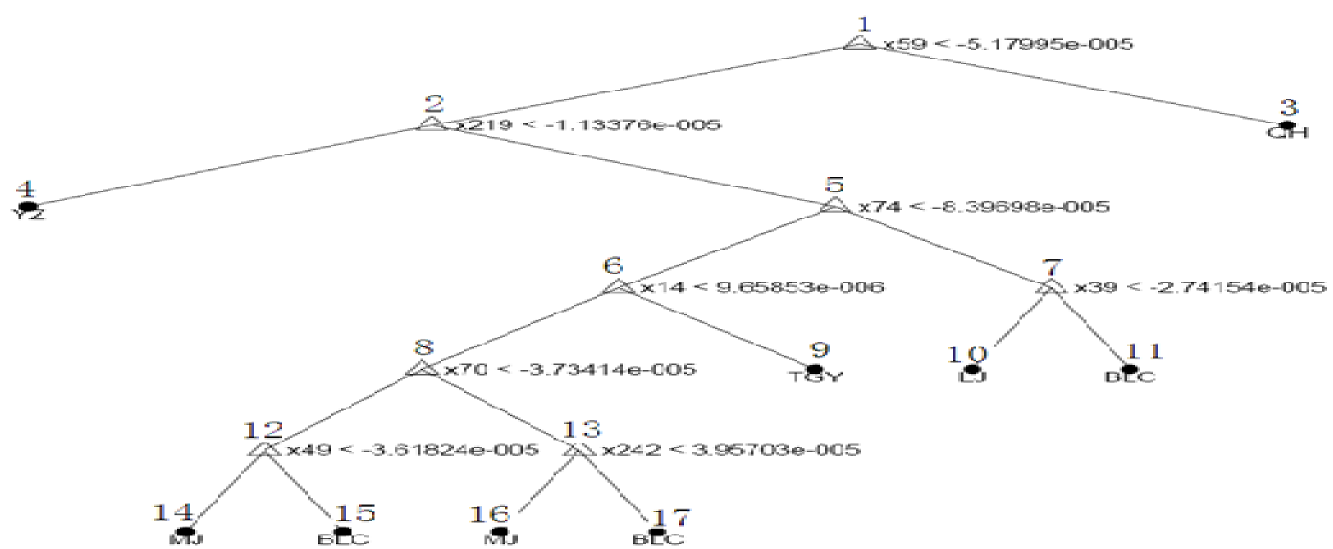

Figure 6. Tree obtained by CART, where △ represents the internal node and ● refers to the leaf node. Each internal node is furnished with the splitting rule, that is, the splitting variable and splitting value. Each leaf node is provided with a node output. Thereinto, BLC, LJ, MJ, QH, TGY, and YZ, respectively, represent Biluochun, Longjing, Maojian, Qihong, Tieguanyin, and Yinzhen tea samples.

both the calibration and prediction sets and slightly lower RRs for Longjing and Tieguanyin in these two subsets. For the varieties of Qihong and Yinzhen, both CART and PLS-DA provided fully accurate discrimination in the calibration and prediction sets. All of these are clearly shown in Table 2, indicating that the inter-variety difference of tea can be well discriminated by using NIR spectroscopy coupled with CART. This may be due to the fact that CART is in possession of high potentials in modeling complex data set, such as fitting strong non-linearity, being immune to heteroscedasticity and suitable for multi-class issue. Moreover, the time required to perform CART is only about two seconds.

Conclusion

In this study, tea variety discrimination, exemplified as Biluochun, Longjing, Maojian, Qihong, Tieguanyin, and Yinzhen, was carried out by using NIR reflectance spectroscopy coupled with pattern recognition techniques. Before the multivariate recognition analysis, data quality control was performed, including outlier and non-linearity detection. Experimental results showed that NIR spectroscopy combined with CART was indicated to hold great potential as an accurate, rapid, and non-invasive strategy for identifying the tea quality.

ACKNOWLEDGEMENTS

The authors are grateful to the national natural science foundation of China (No. 21105035), the self-determined research funds of CCNU from the colleges' basic research and operation of MOE (No. CCNU09A01012), the fundamental research funds for the central universities (Nos. 111016 and 20110348), and the open funds of state key laboratory of chemo/biosensing and chemometrics of Hunan University (No. 200910) for supporting the research work.

REFERENCES

Alain C, Martine D, Marcia V (2008). Nondestructive measurement of fresh tomato lycopene content and other physicochemical characteristics using visible-NIR spectroscopy. J. Agric. Food Chem. 56: 9813-9818.

Breiman L, Friedman JH, Olshen RJ, Stone CJ (1984). In Classification and Regression Trees. Bickel PJ; Cleveland W. S.; Dudley R. M.; Eds.; Wadsworth Internal Group: Belmont, CA.

Centner V, De Noord OE, Massart DL (1998). Detection of nonlinearity in multivariate calibration. Anal. Chim. Acta. 376: 153-168.

Christoffer A, Jonas J, Stefan AE, Sune S, Staffan F (2005). Time-resolved NIR spectroscopy for quantitative analysis of intact pharmaceutical tablets. Anal. Chem. 77: 1055-1059.

Chen QS, Zhao JW, Huang XY, Zhang HD, Liu MH (2006). Simultaneous determination of total polyphenols and caffeine contents of green tea by near-infrared reflectance spectroscopy. Microchem. J. 83: 42-47.

Chen QS, Zhao JW, Liu MH, Cai JR, Liu JH (2008). Determination of total polyphenols content in green tea Using FT-NIR spectroscopy and different PLS algorithms. J. Pharm. Biomed. Anal. 46: 568-573.

Fujiki H, Suganuma M, Okabe S, Sueoka E, Sueoka N, Fujimoto N, Goto Y, Matsuyama S, Imai K, Nakachi K (2001). Cancer prevention with green tea and monitoring by a new biomarker, hnRNP B1. Mutat. Res. 480-481: 299-304.

Hall MN, Robertson A, Scotter CNG (1988). Near-infrared reflectance prediction of quality, theaflavin content and moisture content of black tea. Food Chem. 27: 61-75.

He Y, Li XL, Deng XF (2007). Discrimination of varieties of tea using near-infrared spectroscopy by principal component analysis and BP model. J. Food Eng. 79: 1238-1242.

Jiang JH, James Berry R, Siesler HW, Ozaki Y (2002). Wavelength interval selection in multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data. Anal. Chem. 74: 3555-3565.

Jens B, Jacques W (2008). Industrial applications of online monitoring of drying processes of drug substances using NIR. Org. Process Res. Dev. 12: 235-242.

Jing M, Cai WS, Shao XG (2010). Quantitative determination of the components in corn and tobacco samples by using near-infrared spectroscopy and multiblock partial least s`quares. Anal. Lett. 43: 1910-1921.

Liu YL, Chen YR, Ozaki Y (2000). Two-dimensional visible/near-infrared correlation spectroscopy study of thermal treatment of chicken meats. J. Agric. Food Chem. 48: 901-908.

Luypaert J, Zhang MH, Massart DL (2003). Feasibility study for the use of near-infrared spectroscopy in the qualitative and quantitative analysis of green tea, Camellia Sinensis (L.). Anal. Chim. Acta. 478: 303-312.

Massart DL, Vandeginste BGM, Buydens LMC, Jong SDe, Lewi PJ, Smeyers-Verbeke J (1997). In Handbook of Chemometrics and Qualimetrics: Part A. Elsevier: Amsterdam, the Netherlands

Nakachi K, Matsuyama S, Miyake S, Suganuma M, Imai K (2000). Preventive effects of drinking green tea on cancer and cardiovascular disease: epidemiological evidence for multiple targeting prevention. BioFactors, 13: 49-54.

Schulz H, Engelhardt UH, Wegent A, Drews HH, Lapczynski S (1999). Application of Near-infrared reflectance spectroscopy to the simultaneous prediction of alkaloids and phenolic substances in green tea leaves. J. Agric. Food Chem. 47: 5064-5067.

Setiawan VW, Zhang ZF, Yu GP, Lu QY, Li YL, Lu ML, Wang MR, Guo CH, Yu SZ, Kurtz RC, Hsieh CC (2001). Protective effect of green tea on the risks of chronic gastritis and stomach cancer. Int. J. Cancer, 92: 600-604.

Sun LY, Yang TM, Wang YY (2008). Identification of cortex phellodendri Chinese by near-infrared spectroscopy fingerprint. Comput. Appl. Chem. 25: 329-332.

Shao XG, Bian XH, Cai WS (2010). An improved boosting partial least s quares method for near-infrared dpectroscopic quantitative analysis. Anal. Chim. Acta. 666: 32-37.

Šašić S, Ozaki Y (2001). Short-Wave Near-infrared spectroscopy of biological fluids. 1. quantitative analysis of fat, protein, and lactose in raw milk by partial least-squares regression and band assignment. Anal. Chem. 73: 64-71.

Yang CS, Maliakal P, Meng XF (2002). Inhibition of carcinogenesis by tea. Annu. Rev. Pharmacol. Toxicol. 42: 25-54.

Zhang MH, Luypaert J, Fernández Pierna JA, Xu QS, Massart DL (2004). Determination of total antioxidant capacity in green tea by near-infrared spectroscopy and multivariate calibration. *Talanta,* 62: 25-35.